

Metadatamodel Beeld en Geluid biedt gebruiker van digitale content diepte en structuur

In iMMix, de multimediacatalogus van Beeld en Geluid, kan naar een av-productie op vele manieren worden gekeken, van 'artistiek concept' tot fysieke drager. En dat is ook precies wat de verschillende gebruikers van de collecties willen. Aan dit gloednieuwe catalogussysteem ligt een uitgekiend metadatamodel ten grondslag, dat in hoge mate schatplichtig is aan het FRBR-model van de IFLA.

Annemieke de Jong

Net als het nieuwe gebouw van het Nederlands Instituut voor Beeld en Geluid is iMMix vorig jaar opgeleverd en in gebruik genomen. Aan de bouw van het systeem lagen drie belangrijke ontwikkelingen ten grondslag. Allereerst bood de digitalisering van het omroep-productieproces de mogelijkheid om materiaal binnen een virtueel 'totaalsysteem' te gaan produceren, uit te zenden en te archiveren.* Aan Beeld en Geluid, als partner in het Digitaal Platform – het

samenwerkingsverband met omroepen en NOB dat deze infrastructuur ging opzetten –, de opdracht een archiefcatalogus te bouwen die hier naadloos op aansloot. Een andere aanleiding vormde de digitalisering van de eigen collecties van het instituut. Alleen een modern, eigentijds systeem zou kunnen voldoen aan de zware logistieke, administratieve en technische eisen die bij dergelijke grootschalige processen komen kijken. Ten slotte wilde Beeld en Geluid de digitale dienstverlening uitbreiden. Meer en nieuwe klantgroepen moeten archiefmateriaal online kunnen opvragen en uitwisselen.

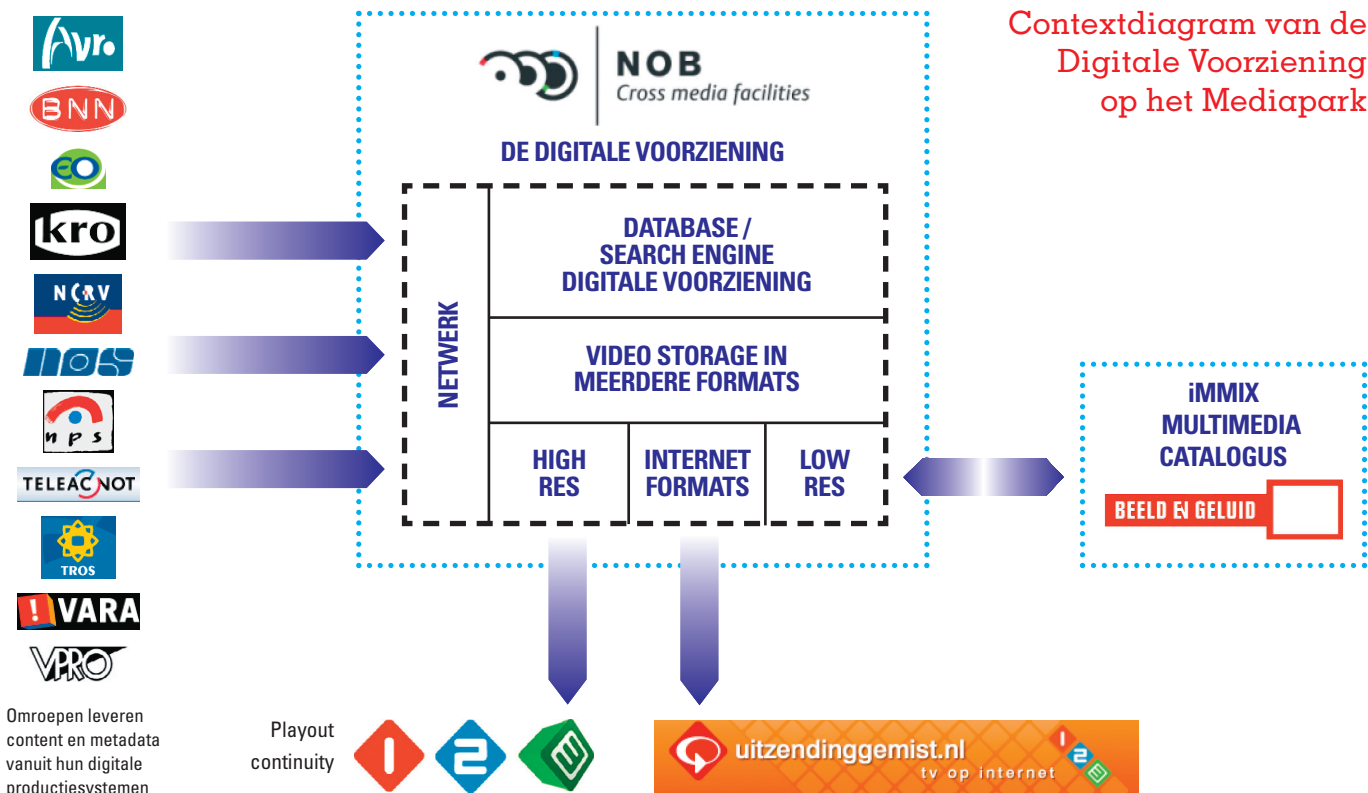
Kijken naar een av-productie

Hoe kun je naar een radio- of televisieprogramma kijken? Welke verschillende 'informatielagen' kun je onderscheiden en hoe is zo'n productie op te delen in losse items? Allereerst is het vaak het product van een omroep, dat eenmalig wordt uitgezonden of hoort bij een serie. Achter zo'n programma zit een idee en in die zin is er sprake van een artistiek of intellectueel concept. Zo'n concept kan op meer manieren worden uitgewerkt en gepubliceerd,

via radio of televisie maar ook in de bioscoop of op internet. Een av-productie kan ook worden beschouwd als een verzameling losse fragmenten, die kunnen worden hergebruikt in een nieuw televisieprogramma, als educatief materiaal in de klas of om thuis achter de computer mee te plakken en te knippen. In fysiek opzicht ten slotte is een programma een tape of – inmiddels vooral – een reeks enen en nullen: een file.

Media Asset Management

De 'Digitale Voorziening', zoals de nieuwe infrastructuur op het Mediapark werd gedoopt, draait inmiddels op volle toeren. Digitaal vervaardigde producties van de omroepen worden vlak voor en na uitzending beheerd in het Digitaal Archief, dat is voorbereid op een jaarlijkse instroom van ongeveer 12 duizend uur video en 20 duizend uur audio. Ook de digitale versies van 'oude' analoge Beeld en Geluid-collecties worden hier opgeslagen. Voor de Media Experience, de publieksattractie, is intussen 13 dui-



zend uur aan hoogtepunten uit de omroepgeschiedenis beschikbaar. Materiaal dat is gedigitaliseerd voor educatief gebruik en (inter)nationale innovatieprojecten is ook in het Digitaal Archief te vinden. In het kader van het project Beelden voor de Toekomst wordt aan deze centrale *storage*-faciliteit de komende zeven jaar nog eens zo'n 150 duizend uur toegevoegd.

Voor de organisatie van de metadata die al deze content beschrijft is er iMMix. iMMix vormt het centrale Media Asset Management-systeem dat programma's vanuit het productieproces importeert, ze aan de hand van 'meegeleverde' metadata indexeert en content en metadata vervolgens beheert en beschikbaar stelt. Het systeem fungeert zo als spin-in-het-web voor iedereen op het Mediapark die werkt met digitale av-content: programmamakers en producenten, editors en technici, documentaristen en collectiemangers.

Voordat maar één code voor het iMMix-systeem was geschreven, had Beeld en Geluid het conceptuele metadatamodel zelf al uitgedokterd. Met dit proces werd begonnen vanuit het inzicht dat de eisen voor een nieuwe catalogus in feite draaien om hetzelfde thema. Of het nu

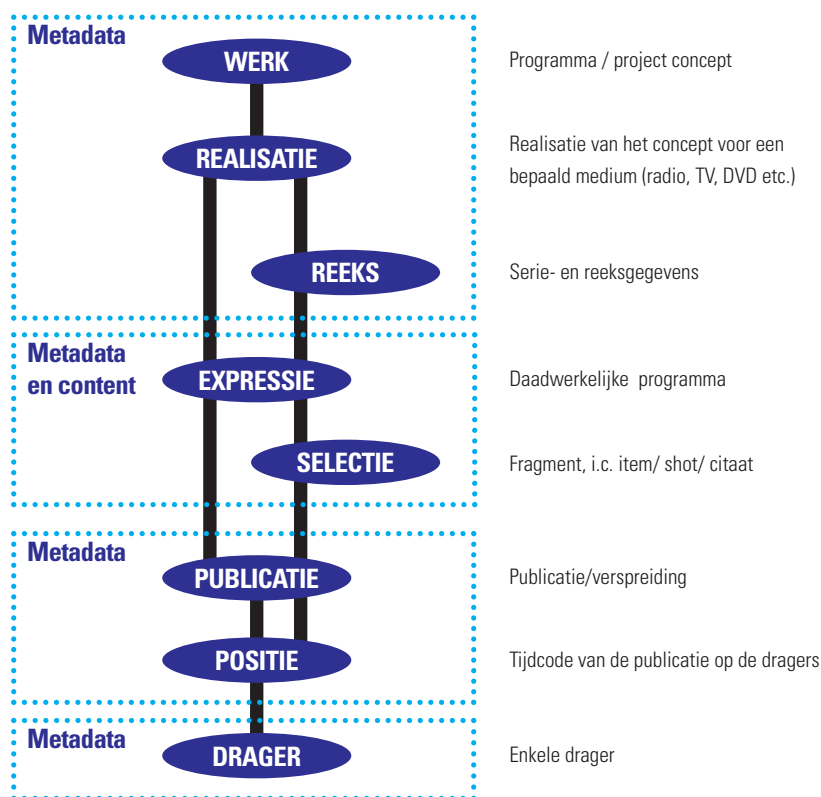
gaat om de beheersing van digitale instroom vanuit de productieomgeving, om het gedifferentieerd kunnen bedienen van online gebruikersgroepen of om het managen van grootschalige digitaliseringsoperaties: vóór alles moet digitale content op meerdere niveaus gekoppeld kunnen worden aan bijbehorende formele, inhoudelijke en 'fysieke' gegevens. De filosofie achter het iMMix-metadatamodel, de conceptuele gegevensarchitectuur van het systeem, werd dan ook gebaseerd op dit uitgangspunt: het gelaagd kunnen identificeren van content en metadata. De groepering van de catalogusgegevens wordt zo een representatie van de wijze waarop de interne en externe gebruikers van Beeld en Geluid de collecties willen benaderen.

Eisen

Feit is dat niet alle gebruikers en raadplegers van audiovisueel materiaal op dezelfde manier kijken. Velen zijn alleen op zoek naar specifieke onderdelen van een programma, zoals een fragment of een citaat, en hebben geen boodschap aan informatie over het hele programma. Anderen zoeken 'abstracter', bijvoorbeeld op gegevens die een serie, genre of

'De gebruikers van de catalogus mogen niet meer of minder gegevens op hun scherm krijgen, dan waaraan ze behoefte hebben'

Schematische weergave van het conceptuele iMMix-metadatamodel



reeks programma's betreffen. Deze klanten willen niet lastig gevallen worden met informatie over de afzonderlijke afleveringen, laat staan over de individuele shots en items. Data over fysieke en administratieve kenmerken en bewerkingen van materialen moet ook weer apart kunnen worden opgevraagd en losstaan van elke beschrijving van de inhoud. En aan de invoerkant, de positie van de documentalisten, moeten deze verschillende facetten van één audiovisuele productie juist als één geïntegreerd, samenhangend geheel kunnen worden benaderd.

Een belangrijke overweging bij het ontwerp van het metadatamodel was dan ook die van de *functionele granulariteit*: de gebruikers mogen niet meer of minder gegevens op hun scherm krijgen, dan waaraan ze behoefte hebben. De eis van de unieke identificatie van alle aangewezen (groepen) metadata en content hing hiermee nauw samen. Ander uitgangspunt vormde de verbetering van de efficiency van het annotatieproces. In de digitale netwerk omgeving moet het nieuwe catalogussysteem voldoen aan het 'write once, read & edit many' principe: gegevens worden in het hele werkproces (dat wil zeggen: de productie-, uitzend- en archiefomgeving) maar één keer ingevoerd. Overerfbaarheid van gegevens is hierbij ook van belang: metadata op hogere beschrijvingsniveaus moet automatisch doorstromen naar de lagere onderdelen. Een laatste eis aan het model was die van de uitbreidbaarheid: te allen tijde moeten gegevens kunnen worden toegevoegd over andere, nieuwe Beeld en Geluid-collecties en -processen.

Het FRBR-datamodel (Functional Requirements for Bibliographical Records) dat in 1998 is ontwikkeld door de Internationale Federatie van Bibliotheek Associaties (IFLA) bleek van aanvang af de grootste inspiratiebron voor het iMMix-metadatamodel. De grondgedachte van dit model, de bekende samenhang tussen de vierdeling 'werk', 'expressie', 'manifestatie' en 'exemplaar' kon goed worden vertaald naar de gewenste ordening van 'audiovisuele' metadata. Voor toepassing binnen het mediadomein waren wel wat uitbreidingen nodig.

Beschrijvingsniveaus

Onder Werk wordt in iMMix verstaan:

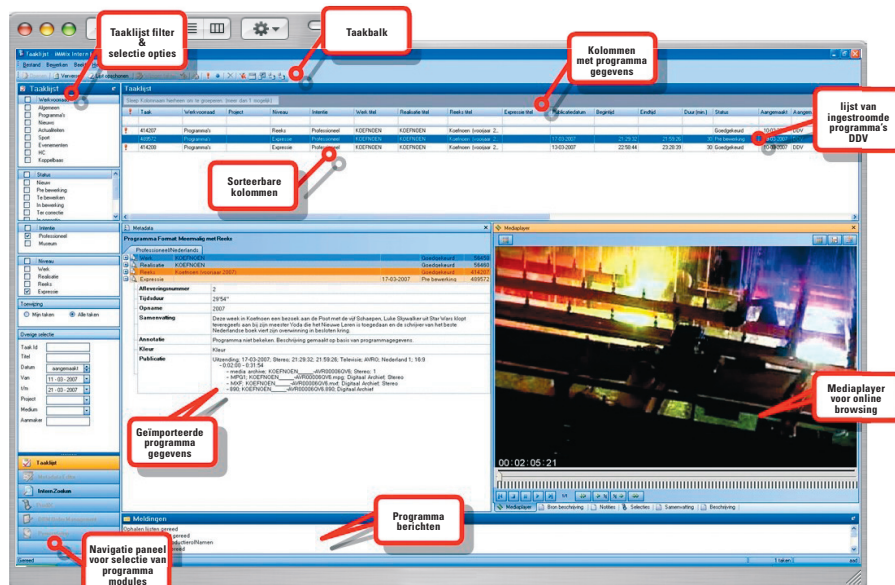
Contextualisering

iMMix kent een zogenaamde 'intentieconstructie'. Intenties zijn versies van een beschrijving. Als eenmaal een catalogusbeschrijving is gemaakt, kan aan het systeem een andere versie worden toegevoegd, waarin een beperkt aantal velden op een afwijkende manier wordt ingevuld. Deze versie wordt gekoppeld aan de bestaande beschrijving en is verbonden aan dezelfde digitale content. Intenties creëren de mogelijkheid domeinspecifieke en/of additionele gegevens aan de eerstgemaakte beschrijving te hangen en deze gegevens autonoom in te voeren en te presenteren. Doel van het maken van intenties is het kunnen inspelen op specifieke informatiebehoeften van verschillende gebruikersgroepen. Voor de toepassing kan, naast het professionele av-domein, wor-

den gedacht aan gebruik door het algemeen publiek (bijvoorbeeld de beschrijvingen van archiefmateriaal dat wordt gepresenteerd in de Media Experience). Zo zal het veld 'samenvatting' in een publiekbeschrijving van een programma van Paul de Leeuw informatie bevatten over de populariteit, het maatschappelijk effect en mogelijke televisieprizen die zijn gewonnen. In de beschrijving voor de professionele gebruikers is de samenvatting 'objectiever' en meer gericht op de inhoud van het betreffende programma zelf. Er kunnen ook intenties worden gemaakt ten behoeve van het educatieve veld. Hierin kunnen bijvoorbeeld trefwoordvelden worden ingevuld aan de hand van domeingebonden thesauri. Gebruik van intenties wordt inmiddels aangeduid als 'contextualisering'.

het artistieke en intellectuele concept, het idee achter een productie. Het werkniveau fungeert als een kapstok waaraan meerdere Realisaties kunnen zijn opgehangen. Realisaties zijn uitwerkingen van een Werk, bestemd voor een bepaald medium (televisie, radio, internet, dvd etcetera). Een voorbeeld is het Werk 'Soldaat van Oranje', waaraan de Realisaties 'serie' en 'speelfilm' kunnen worden gehangen. Het Werk 'NOS journaal' kan bijvoorbeeld uiteenvallen in de Realisaties 'Jaaroverzicht', 'Jeugdjournaal' en '8-uur journaal'. Een volgend (optioneel) niveau groepeert de gegevens over de reeks of de serie. Van de Realisatie 'serie Soldaat van Oranje' worden de seriegegevens op Reeksniveau vastgelegd. De Realisatie 'speelfilm Soldaat van Oranje' kent dit niveau niet. Het Expressieniveau vervolgens, is de concrete uitvoering van een Realisatie. Hier gaat het om metadata over de daadwerkelijke productie, de film Soldaat van Oranje, of een van de afleveringen van de gelijknamige serie. Een Expressie kan Selecties bevatten, dat wil zeggen apart geïdentificeerde (afgemonteerde) items en fragmenten of citaten.

In het metadatamodel van iMMix wordt het manifestatieniveau van het FRBR-model Publicatie genoemd. Een Expressie kan worden uitgebracht en verspreid via verschillende distributiekanaalen (televisie, radio, internet). Terwijl het bij het niveau Realisatie gaat om het medium dat werd *beoogd*, legt Publicatie de *feitelijke* distributievorm vast van een aangewezen Expressie. Publicatie van een programma kan immers uiteindelijk (ook) via een heel ander medium plaatsvinden, een veelvoorkomend verschijnsel in deze tijd van convergentie (bijvoorbeeld een documentaire die oorspronkelijk is gemaakt voor televisie en die ook via het Internet wordt uitgezonden). Het exemplaarniveau van het FRBR-model vervolgens, valt in iMMix uiteen in Positie (de tijdcode van de Publicatie van de Expressie en Selecties op de digitale en/of analoge dragers), met daaronder het niveau Drager (gegevens over de individuele file, tape of band). Het metadatamodel positioneert de gegevens op zo'n manier dat de aangewezen groepen informatie apart te benaderen zijn, maar onderling wel gekoppeld blijven. Aldus verwijst een deel naar het geheel, de serie naar de aflevering, de collectie naar het onderdeel, het programma



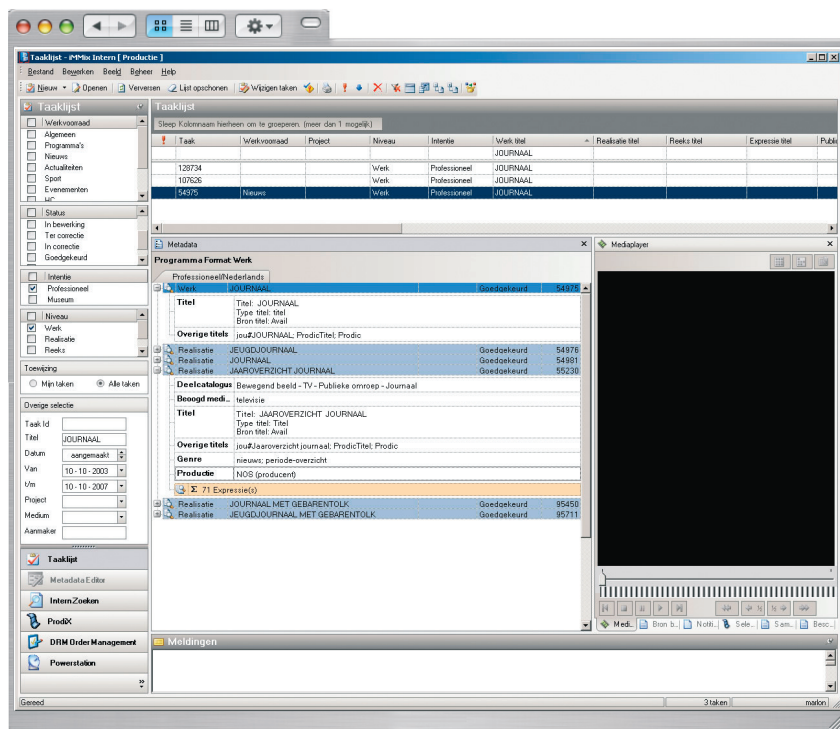
het iMMix- 'dashboard' van de documentalist

naar de uitzending, de fysieke drager naar de inhoud en vice versa. De verschillende catalogusobjecten die horen bij één overkoepelend object worden op deze manier samengebracht.

Attributen

Aan alle beschrijvingniveaus van het model worden attributen of metadatatelden gehangen. Koppeling met de digitale content is er alleen op Expressie- en Selectieniveau – het daadwerkelijke programma. Er zijn attributen die niveau-specifiek zijn (bijvoorbeeld gegevens over de Reeks en over de Drager) en er zijn attributen die op meer niveaus voorkomen, waaronder een aantal beschrijvende velden en de termen uit de GTAA (Gemeenschappelijke Thesaurus Audiovisuele Archieven). Het Werkniveau, de kapstok waaraan alle manifestaties van een bepaald audiovisueel concept worden opgehangen, kent geen specifieke metadata. Hier wordt slechts gesproken van Uniforme Titel. Voor de definitie van de attributen is een eigen metadata-dictionary samengesteld, die werd gedestilleerd uit alle descriptorren die ooit werden gebruikt in de oude catalogi en databases van Beeld en Geluid. Dit extract werd 'gemapt' met een aantal internationale standaarden, zoals de Minimal Data List van de International Federation of Television Archives FIAT/IFTA en uitwisselingsstandaarden voor de omroepwereld,

'Het Digitaal Archief is voorbereid op een jaarlijkse instroom van ongeveer 12 duizend uur video en 20 duizend uur audio'



De verschillende Realisaties van het Werk 'NOS Journaal'

waaronder de door de European Broadcasting Union ontwikkelde standaard P/Meta. Voor export naar Dublin Core werd een XML-protocol opgesteld.

Naast de metadata die horen bij de recent uitgezonden programma's bevat de iMMix-catalogus intussen ook de vele honderdduizenden beschrijvingen uit de oude catalogi en bestanden van Beeld en Geluid. Deze metadata verwijst (nog) vooral naar analoge dragers. Om de veelal ongestructureerde beschrijvingen – afkomstig van de fusiepartners van Beeld en Geluid, de Stichting Film- en Wetenschap, het Filmarchief van de RVD en het Audiovisueel Archiefcentrum – op een ordentelijke manier te laten instromen in de gelaagde opdeling binnen de iMMix metadatastructuur, waren vele ingenieuze conversiescripts nodig.

Het metadatamodel voor iMMix is in eerste instantie bedacht om het (actuele) digitale beeld- en geluidsmateriaal van de omroepen te kunnen opslaan, verwerken en aanbieden. Voor het ontsluiten van de muziekcollectie van Beeld en Geluid – de platen en cd's gemaakt voor de handel – bleek het model in deze vorm minder bruikbaar. Een audiovisuele (omroep-)productie is uniek en valt in verreweg de meeste gevallen samen met zijn expressie. Bij muziekwerken is dit niet het geval. Voor de bouw van de muziekcatalogus, iMMix-M, moest het FRBR-model daarom enigszins op zijn kop worden gezet. Uitgangsniveau is hier veeleer de Drager, die Expressies (hier: de specifieke uitvoeringen) van een bepaalde Werk (hier: een compositie) kan bevatten. Een Drager kan zelf ook een Werk zijn. Dit geldt bijvoorbeeld bij verzamel-cd's.

Voor een aantal historische collecties van Beeld en Geluid (met name de heterogene materialen die niet of niet in hun geheel voor uitzending bestemd zijn geweest) wordt het model momenteel uitgebreid met een extra hiërarchische tak. Hoogste niveau vormt hier de 'collectie' van een bepaalde maker, verzamelaar of depotgever. Zo'n 'collectie' kan – naast films en tapes – ook boeken, papieren documenten en objecten omvatten. De nevenstructuur voor gegevens over deze materialen mondt uit in het Dragerniveau van het hoofdmodel en kent, voorzover het de audiovisuele onderdelen betreft, diverse dwarsverbanden met de andere niveaus.

Programmaformats

Het iMMix-metadatamodel is een neerslag van het *totaal* aan mogelijke metadatavelden en mogelijke informatieniveaus. Voor de ontsluiting van niet elke productie [Deze zin loopt niet] zijn alle velden en niveaus noodzakelijk. Er zijn programma's die niet behoren tot een Reeks en programma's waarbij het niet zinvol is ze te ontleden in shots of items (Selecties). Ook de toekenning van beschrijvende metadata verschilt per programmasoort. Om vast te leggen welk soort programma met welke niveau en welke metadatavelden ontsloten moet worden is het concept van de 'programmaformats' bedacht. Programmaformats zijn maskers of 'views' op het totaal aan velden en

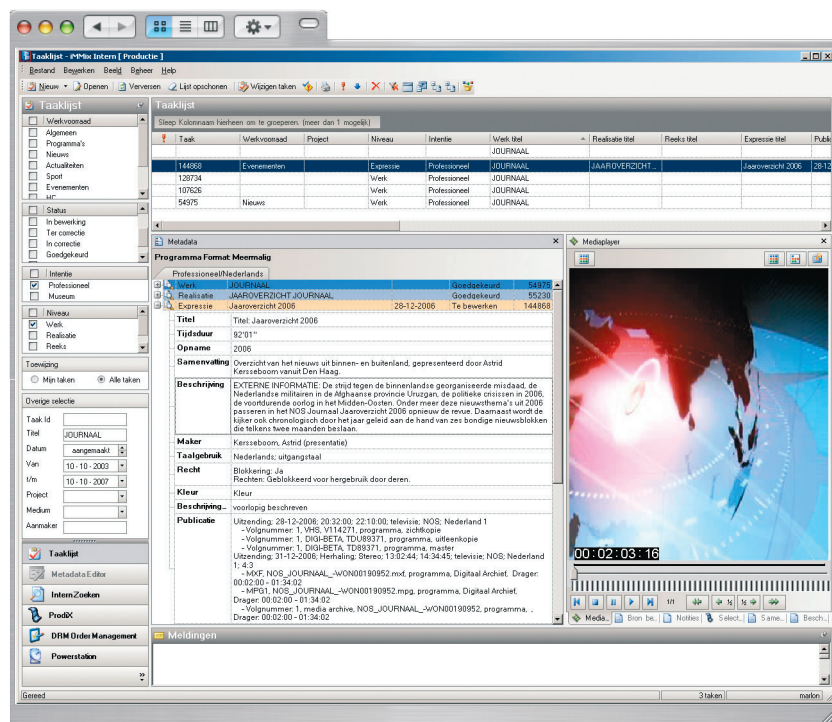
niveaus. Deze views zijn voor wat betreft de niveaus samengesteld op grond van de structuur (deel-geheelverhouding) en voor wat betreft de hoeveelheid metadatavelden op de gewenste diepte van ontsluiting. Er zijn twee redenen om een programmaformat aan te maken. Het kan gaan om afwijkend gestructureerd materiaal binnen dezelfde collectie (in het bijzonder voor wat betreft de verhouding deel-geheel). Hier verschilt vooral de keuze van informatieniveaus binnen het format. Het kan ook gaan om een andersoortige collectie zoals de RTV-collectie, de fotocollectie of de internetcollectie. Hier verschilt vooral het soort metadatavelden.

Het model in de praktijk

In de netwerkomgeving is een gedifferentieerde opdeling een keiharde voorwaarde voor het doelmatig opslaan, managen en distribueren van digitale av-content en beschrijvingen, zeker als het gaat om hoeveelheden zoals die dagelijks circuleren op het Mediapark. Voor Beeld en Geluid, haar systemen en haar eindgebruikers biedt de implementatie van het iMMix-metadatamodel dan ook grote voordelen. Als de verschillende objecten (programma's, items, shots, series, citaten etcetera) niet apart zouden kunnen worden geïdentificeerd, kunnen ze bijvoorbeeld niet verwijzen naar een tijdcode in de content en zouden complete programma's moeten worden afgespeeld om bij het gevraagde onderdeel terecht te komen. Het zou dan ook niet mogelijk zijn om auteursrechten te koppelen aan specifieke shots of fragmenten, een belangrijke voorwaarde voor het beschermen van rechten in een online omgeving. De interne en externe workflows kunnen door zo'n opdeling in losse objecten daarbij veel beter worden beheerst en gecontroleerd. Een ander voordeel betreft de gebruiksrechten. Door het opknippen van content en metadata kan de toegang per interne en externe klantgroep verfijnd worden ingeregeld. Wat ook steeds meer aan belang toeneemt is het op fragment en shotniveau kunnen koppelen van de software voor automatische indexerend, zoals die momenteel wordt ontwikkeld. Voor de ruim vijftien documentaristen van Beeld en Geluid is de overgang van een ongestructureerde 'stand alone' catalogusomgeving naar een groot gereguleerd informatienetwerk redelijk ingrijpend. Zij worden tijdens hun dagelijks werk voortdurend geconfronteerd met de strenge en soms weerbarstige hiërarchie van het model. Ze moeten leren denken in losse catalogusobjecten en hebben voor de ontsluiting van één integraal programma méér invoerstappen nodig, waarbij per beschrijvingsniveau ook nog eens specifieke regels gelden. Hun rollen en rechten bij de aanmaak, bewerking, mutatie en goedkeuring van de documenten zijn scherp afgebakend. Omdat een groot deel van de metadata van de dagelijkse aanwas aan radio- en televisieprogramma's wordt geïmporteerd uit een andere omgeving, is momenteel ook nog

veel inspanning nodig voor correctie en controle achteraf. Efficiencywinst bij het annoteren is daarbij niet onmiddellijk zichtbaar, omdat van elk nieuw ingestroomd programma eerst de hoger gelegen Werk-, Realisatie- en Reeksniveaus moeten worden aangemaakt.

De gewijzigde structuren voor workflow en catalogisering openen echter ook nieuwe arbeidsperspectieven. Doordat veel (formele) metadata automatisch instroomt, zullen av-documentaristen bij Beeld en Geluid zich uiteindelijk vooral kunnen gaan bezighouden met het vrijmaken en verdiepen van de inhoudelijke beschrijvingen. Ook het thematiseren en contextualiseren van zowel materiaal uit de nieuwe als de oude digitale collecties kan een nieuwe taak gaan vormen. Het verder ontwikkelen en uitdragen van de eigen metadataconventies wordt ook steeds belangrijker voor de consistentie en integriteit van informatie in de netwerkomgeving. Hierbij hoort bijvoorbeeld het propageren van geüniformeerde invoerregels en het gebruik van de Gemeenschappelijke Thesaurus Audiovisuele Archieven (GTAA) door alle gebruikers binnen de 'Digitale Voorziening'. <



D. De Expositie 'Jaaroverzicht 2006' van de Realisatie 'Jaaroverzicht NOS Journaal'

‘De filosofie achter het iMMix-metadatamodel is het gelaagd kunnen identificeren van content en metadata’

* Zie ook Kouwenhoven, T., J. Oomen: 'Informatiesysteem brengt mogelijkheden digitalisering binnen bereik. Bladeren door Beeld en Geluid'. InformatieProfessional, 3/2006.

Drs. A.S.M. de Jong is Hoofd Informatiebeleid bij het Nederlands Instituut voor Beeld en Geluid en verantwoordelijk voor de ontwikkeling van het conceptuele metadatamodel voor het iMMix-systeem.